



# Directed Evolution of a Small-Molecule-Triggered Intein with Improved Splicing Properties in Mammalian Cells

## Citation

Peck, Sun H., Irwin Chen, and David R. Liu. 2011. "Directed Evolution of a Small-Molecule-Triggered Intein with Improved Splicing Properties in Mammalian Cells." *Chemistry & Biology* 18 (5) (May): 619–630. doi:10.1016/j.chembiol.2011.02.014.

## Published Version

doi:10.1016/j.chembiol.2011.02.014

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:25271391>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

*Chem Biol.* 2011 May 27; 18(5): 619–630. doi:10.1016/j.chembiol.2011.02.014.

## Directed Evolution of a Small Molecule-Triggered Intein with Improved Splicing Properties in Mammalian Cells

Sun H. Peck, Irwin Chen, and David R. Liu\*

Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138 USA

### Summary

Laboratory-created small-molecule-dependent inteins enable protein structure and function to be controlled post-translationally in living cells. Previously we evolved two inteins (2-4 and 3-2) that splice efficiently in the presence, but not the absence, of the cell-permeable small molecule 4-hydroxytamoxifen (4-HT) in a variety of extein contexts in *Saccharomyces cerevisiae*. In mammalian cells, however, the 2-4 and 3-2 inteins exhibited significantly lower splicing efficiencies and slower splicing in the presence of 4-HT, as well as higher background splicing in the absence of 4-HT, than in yeast cells. In this work we evolved the 2-4 and 3-2 inteins through several additional rounds of mutation, recombination, and screening in *S. cerevisiae* at both 30 °C and 37 °C. The resulting second-generation evolved inteins exhibit substantially improved (~2- to 5-fold higher) splicing yields in yeast compared to the parental 2-4 and 3-2 inteins and significantly faster splicing kinetics. The improved properties of these evolved inteins carried over to mammalian cells, in which the newly evolved inteins spliced with substantially greater (~2- to 8-fold) efficiency in the presence of 4-HT while maintaining background splicing levels in the absence of 4-HT that are comparable to or better than the levels observed with the 2-4 or 3-2 inteins. In total, these inteins were tested in four different protein contexts in yeast and human cells and found to exhibit their substantially improved properties in all contexts tested, typically resulting in 50–90% spliced protein in the presence of 4-HT and < 5% splicing in the absence of 4-HT. The second-generation evolved inteins augment the promise of ligand-dependent protein splicing as an effective and broadly applicable approach to probing protein function in mammalian cells.

### Introduction

Methods to control protein structure and function inside living cells have proven to be valuable tools to elucidate the roles of proteins in their native biological contexts (Schreiber, 2003; Buskirk and Liu, 2005; Banaszynski and Wandless, 2006). Traditional genetic methods that have been widely used to control protein function by altering expression levels in mammalian cells include knock-out and knock-in systems such as those mediated by Cre-Lox recombination (Sauer et al., 1988) and the use of transcriptional regulators such as the tetracycline-responsive tet-on/tet-off systems (Gossen et al., 1992). These methods are highly specific to the protein of interest and can be applied to many proteins, but typically

© 2011 Elsevier Ltd. All rights reserved.

\*Professor David R. Liu, Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138 USA, drliu@fas.harvard.edu, (617) 496-1067 (office), (617) 496-5688 (fax).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

require days to reach steady-state protein levels in mammalian cells, are irreversible in the case of recombination-based methods, and are vulnerable to transcriptional compensation (Shogren-Knaak et al., 2001; Marschang et al., 2004; Wong and Roth, 2005; Acar et al., 2010). Other methods such as RNA interference (Fire et al., 1998), chemical genetics (Specht and Shokat, 2002), small-molecule regulated protein stability or degradation (Stankunas et al., 2003; Schneekloth et al., 2004; Banaszynski et al., 2006), and small-molecule induced proteolytic shunts (Pratt et al., 2007) have also been used effectively by many researchers and offer more rapid control over protein levels than strategies that exert control before transcription, but can require the discovery of small-molecule modulators of protein function, necessitate the involvement of other cellular machinery that may not be present in the cells of interest, or are prone to off-target effects.

Protein splicing elements known as inteins are able to catalyze their excision out of a single polypeptide and leave behind the flanking sequences, or exteins, precisely ligated together through a native peptide bond (Paulus, 2000). Inteins are attractive choices as tools for modulating protein expression because they do not require any other cellular components, are able to splice out of a wide variety of extein contexts (Xu et al., 1993), and can undergo splicing in minutes (Paulus, 2000). Although natural inteins splice spontaneously, inteins that undergo splicing in a small-molecule-dependent manner have been developed by fusing intein halves with proteins that dimerize in the presence of a small molecule (Mootz and Muir, 2002; Mootz et al., 2003; Shi and Muir, 2005), or by directed evolution in which a library of intact inteins fused to a ligand-binding domain was screened to splice in the presence, but not the absence, of a small molecule (Buskirk et al., 2004). These small-molecule-dependent inteins have enabled protein function in cells to be controlled post-translationally by the addition of an exogenous, cell-permeable molecule (Mootz and Muir, 2002; Mootz et al., 2003; Buskirk et al., 2004; Mootz et al., 2004; Shi and Muir, 2005; Yuen et al., 2006; Schwartz et al., 2007; Hartley and Madhani, 2009).

Previously we developed variants of the *Mycobacterium tuberculosis* RecA intein that selectively splice in the presence of the cell-permeable small molecule 4-hydroxytamoxifen (4-HT) in a rapid, dose-dependent manner using directed evolution in *S. cerevisiae* (Buskirk et al., 2004). The *M. tuberculosis* RecA intein was chosen because it can efficiently splice in a wide variety of contexts (Lew and Paulus, 2002), and the evolved 4-HT-triggered inteins retained this characteristic. These evolved inteins have been successfully used as a tool to study the role of histone H2A.Z in establishing chromatin architecture around promoter regions in *S. cerevisiae* (Hartley and Madhani, 2009).

We showed that these evolved inteins are functional in mammalian cells at 37 °C but splice with significantly reduced speed, lower efficiency, and higher background splicing in the absence of 4-HT compared with splicing at 30 °C in yeast (Yuen et al., 2006). These limitations constrain the utility of these evolved inteins as tools for mammalian cell biology; indeed, only two studies (Mootz et al., 2003; Yuen et al., 2006) have reported the use of small molecule-dependent inteins in mammalian cells. We therefore sought to develop improved intein variants that can splice efficiently, rapidly, and in a highly ligand-dependent manner in mammalian cells at 37 °C. In this work we describe the results of new directed evolution efforts to improve the splicing characteristics of 4-HT dependent inteins for use at 37 °C and in mammalian cells. The resulting inteins in yeast cells exhibit substantially improved splicing activity and speed with no significant increase in background splicing at both 30 °C and 37 °C. These second-generation inteins also splice with much greater speed and efficiency in human cells at 37 °C in four different protein contexts compared with the parental inteins. These new ligand-dependent inteins represent more effective and broadly applicable tools for the small-molecule-triggered, post-translational modulation of protein activities in living systems including mammalian cells.

## Results

### Evolution scheme for improved 4-HT-dependent inteins

To improve the splicing characteristics of the evolved 4-HT dependent inteins, we modified the high-throughput fluorescence-activated cell sorting (FACS) screen that we previously used to isolate active and inactive inteins from mixed starting populations (Buskirk et al., 2004) (Figure 1A). We genetically inserted the 4-HT-dependent intein in place of Cys 108 of GFP(uv), a FACS-optimized GFP mutant, which places the intein near the mid-point of a  $\beta$ -strand and abolishes fluorescence until splicing takes place (Ormo et al., 1996; Buskirk et al., 2004). During positive screens for intein splicing activity, we collected cells that exhibited GFP fluorescence in the presence of 4-HT, while during negative screens we collected cells that remained non-fluorescent in the absence of 4-HT (Figure 1B). Using error-prone PCR with mutagenic dNTPs (Zaccolo et al., 1996), point mutations were randomly introduced into the genes of the two best inteins (including the ligand-binding domain) resulting from our first intein evolution effort, the 2-4 and 3-2 inteins. The resulting intein gene library was cloned into the p416Gal1 vector in *S. cerevisiae* RDY98 using gap repair homologous recombination (Raymond et al., 1999) to obtain a starting library size of  $7 \times 10^6$  clones. This starting library was subjected to two evolution efforts in parallel, one conducted at 30 °C and one at 37 °C (Figure 1C).

Each round of evolution consisted of at least two positive screens and one negative screen (Figure 1B). Positive screen 1 (P1) for each round collected the 5% most fluorescent library members in the presence of 4-HT. The second positive screen (P2) in each round collected library members that exhibited better splicing activity than the parental 3-2 intein in the presence of 4-HT by collecting cells that were more fluorescent than cells transformed with a 3-2 intein-GFP construct. In Round 2, a third positive screen was carried out (P3) that further enriched for library members with better splicing activity than the 3-2 intein in the presence of 4-HT. As the final screening step in each round of evolution, a single negative screen (N) collected library members that did not generate spliced GFP (*i.e.*, were not fluorescent) in the absence of 4-HT. Surviving gene pools were diversified after each round. Following Round 1, the genes of the surviving library members in each of the two libraries were separately mutagenized using error-prone PCR before re-cloning into yeast as the starting library for Round 2. After Round 2, surviving genes from the 30 °C and 37 °C screens were combined and subjected to *in vitro* homologous recombination using the StEP method (Zhao and Zha, 2006). The resulting recombined library was subjected to separate Round 3 screens at 30 °C and at 37 °C. Overall, the entire evolution process comprised three complete rounds containing 10 individual screening steps each for the 30 °C and the 37 °C efforts (Figure 1C).

### Splicing Characteristics of Evolved Inteins

Clones from the 30 °C and the 37 °C libraries surviving each of the rounds of evolution were isolated and the genes encoding their inteins were sequenced. Three sequences each from the 30 °C and 37 °C libraries following Round 3 were selected for detailed characterization on the basis of their high degree of abundance in the final evolved pools. These six intein sequences are summarized in Table 1. The newly evolved clones are designated 30RX-Y (from evolution at 30 °C) or 37RX-Y (from evolution at 37 °C), where X refers to the round number from which the clone was isolated and Y refers to the clone number within that round. Mutations Val34Ala, Ile66Thr, Thr328Lys, and Glu375Gly are shared among clones in both the 30 °C and 37 °C libraries. Leu124Pro was observed only in the 30 °C library, and Asp129Asn and Cys178Arg are only observed in the 37 °C library.

All six evolved inteins were assayed for splicing function in the GFP context in yeast cells at both 37 °C and 30 °C, and their activities were compared with those of the original 2-4 and 3-2 inteins under the same conditions. We appended a FLAG-tag at the C-terminal end of the GFP-intein sequence to facilitate detection of both the spliced and unspliced protein products. Cells treated with 1  $\mu$ M 4-HT or without 4-HT at time points from 1 hour to 24 hours were subjected to FACS analysis in order to detect accumulation of functional GFP after the splicing reaction (Figure S1), and the spliced and unspliced proteins in the corresponding cell lysates were quantified by Western blot (Figure 2A) and densitometry. The percentage of spliced protein was calculated as the amount of spliced protein divided by the total amount of spliced + unspliced protein for each sample.

In yeast cells at 37 °C, all six newly evolved inteins exhibit substantially faster production of spliced proteins as well as a significantly higher percentage of spliced product in the presence of 4-HT as compared to the parental 2-4 and 3-2 inteins (Figure 2B). Compared to the original 2-4 or 3-2 inteins, the most active inteins at 37 °C generate up to 8-fold more spliced protein 3 hours after 4-HT addition and up to 5-fold more spliced protein at 6 hours after 4-HT addition (Figure 2B). For example, while the 2-4 and 3-2 inteins generated 17–27% spliced GFP after 6 hours at 37 °C, the six newly evolved inteins resulted in 70–86% spliced GFP at the same time point. Likewise, while the 2-4 and 3-2 inteins did not generate significant amounts of spliced protein 1 hour after 4-HT treatment, the newly evolved inteins produced 25–35% spliced protein at this early time point. For five of the six clones (all but clone 30R3-3), splicing in the absence of 4-HT remained low (typically < 10% after 24 hours; see Figure 2C). Total protein expression per yeast cell of all six inteins at each time point were within 10% of the expression levels observed with the 3-2 intein (unpublished data), suggesting that the newly evolved inteins do not alter protein expression levels compared with the 3-2 intein and are not unusually susceptible to degradation.

In yeast cells grown at 30 °C, five of the six newly evolved inteins (all but clone 37R3-1) exhibit more efficient splicing by the six-hour time point compared to the 2-4 and 3-2 inteins (Figure 2D). Splicing in the absence of 4-HT after 24 hours was generally  $\leq$  15%, the level of background splicing observed for the 3-2 intein after 24 hours (Figure 2E). The newly evolved inteins generated 3.6- to 7-fold higher levels of spliced protein at 3 hours, and 1.6- to 2.6-fold higher levels of spliced protein at 6 hours, relative to the 3-2 or 2-4 inteins, respectively. At later time points (12 or 24 hours) splicing efficiencies for the newly evolved inteins excluding 37R3-1 were generally high ( $\geq$  70%), similar to that of the 3-2 intein. As observed in the 37 °C assays, the protein expression levels per yeast cell of each intein were within  $\pm$ 10% of the level observed with the 3-2 intein at each of the time points. Taking into consideration splicing rate, overall splicing efficiency in the presence of 4-HT, and background splicing in the absence of 4-HT, the best performing evolved intein clone for use at 37 °C was 37R3-2, and the best clone at 30 °C was 30R3-1. Together, these results indicate that the evolution strategy described above resulted in inteins with substantially improved splicing speed and yields of spliced protein in yeast at 30 °C and especially at 37 °C, without significantly impairing 4-HT dependence.

### Ligand-Dependent Splicing of Newly Evolved Inteins in Mammalian Cells

All six evolved intein sequences described above in the GFP context were cloned into a pCMV promoter-based mammalian expression vector with a C-terminal FLAG-tag for Western blot analysis. HEK293 cells were transfected with these vectors, incubated for 24 hours at 37 °C, then treated with Dulbecco's modified Eagle medium (DMEM):F12 with 10% fetal bovine serum (FBS) containing 1  $\mu$ M final concentration of 4-HT or with the same medium lacking 4-HT. The cells were incubated at 37 °C for an additional 12–24 hours then harvested for qualitative assessment of functional GFP accumulation by FACS (Figure S2) and quantification by Western blot analyses (Figure 3A). Consistent with the

characteristics of the newly evolved inteins in yeast cells, the three assayed clones evolved at 37 °C all exhibit faster GFP splicing kinetics and higher overall splicing yields at both the 12-hour and 24-hour time points compared with the 2-4 or 3-2 inteins (Figure 3B). The best 37 °C clone, 37R3-2, exhibited 3.8-fold and 2.2-fold higher GFP splicing efficiency after 24 hours than the original 2-4 and 3-2 intein, respectively (73% spliced GFP for 37R3-2 vs. 19% for 2-4 and 33% for 3-2). Background splicing in the absence of 4-HT was not observed for 37R3-2 (Figure 3B), further consistent with the high ratio of ligand-induced splicing to background splicing of 37R3-2 observed in yeast cells. The inteins evolved at 30 °C also exhibit similarly improved (~2- to 6-fold) splicing kinetics and splicing efficiencies at 12 and 24 hours relative to that of inteins 2-4 and 3-2 (Figure 3C). The best 30 °C library clone (30R3-1) generated 72% spliced GFP after 24 hours, compared with 33% and 19% spliced GFP for the 3-2 and 2-4 inteins, respectively, while splicing with  $\leq 3\%$  efficiency in the absence of 4-HT (Figure 3C).

Interestingly, the fraction of protein splicing that was completed by 12 hours relative to an endpoint of 24 hours was also greater for the newly evolved inteins compared with the 2-4 and 3-2 inteins. For example, an average of 59% of the total amount of spliced GFP in mammalian cells after 24 hours was present after 12 hours among the three 37 °C library clones, and an average of 57% of the total amount of spliced GFP after 24 hours was present after 12 hours among the three 30 °C library clones. For comparison, 45% and 31% of the total spliced GFP after 24 hours was present at 12 hours for the 3-2 and 2-4 inteins, respectively. These results collectively indicate that in live mammalian cells at 37 °C the newly evolved inteins exhibit increased splicing rate and higher extent of splicing compared with the original evolved inteins, while maintaining low background splicing in the absence of 4-HT.

### Evolved Intein Properties in Different Proteins in Mammalian Cells

Next we studied the generality of 4-HT-dependent splicing of the two best evolved inteins (30R3-1 and 37R3-2) by inserting these inteins into three other protein contexts in addition to GFP in mammalian cells. The inteins were inserted into mCherry, a red fluorescent protein, in place of Thr 113, the residue in mCherry that corresponds to the Cys residue used for intein insertion in GFP. As in the case of GFP, this placement positions the intein near the mid-point of a  $\beta$ -strand and abolishes mCherry fluorescence until splicing takes place (*vide infra*). The corresponding genes were introduced into HEK293 cells as described above for the GFP-intein genes, then treated with media containing 1  $\mu$ M 4-HT or with media lacking 4-HT. These cells were incubated at 37 °C for an additional 12 to 24 hours, then harvested for detection of functional mCherry accumulation by FACS (Figure S3) and quantification by Western blot analysis.

Both of the 37R3-2 and the 30R3-1 evolved inteins continued to exhibit significant improvement in splicing performance over the 2-4 and 3-2 inteins in the context of mCherry (Figure 4A). The 37R3-2 intein resulted in 72% spliced mCherry protein after 24 hours and 43% spliced protein after 12 hours. The 30R3-1 intein resulted in 54% and 33% spliced mCherry after 24 and 12 hours, respectively. Thus the percentage of spliced mCherry generated by the 37R3-2 intein at 24 hours or 12 hours was ~3-fold higher than that of 3-2 intein, and ~5-fold higher than that of the 2-4 intein. Background splicing in the absence of 4-HT was  $\leq 3\%$  for all inteins assayed in this context.

We evaluated the splicing characteristics of the newly evolved inteins in the contexts of two additional mammalian proteins, Gli1 and Gli3T. Gli1 and Gli3 are transcription factors that mediate Hedgehog signaling (Koebernick and Pieler, 2002) and are important in many key developmental processes such as spinal cord patterning (Bai et al., 2004) and limb development (Barna et al., 2005). Gli3T is a C-terminally truncated form of the transcription



factor Gli3 that is used as a transcriptional repressor (Wang et al., 2000). Gli1 and Gli3T are large proteins, 122kDa and 85kDa respectively, and are both structurally unrelated to GFP, mCherry, and structurally distinct from each other. The 37R3-2 and 30R3-1 inteins were inserted genetically in place of Cys 273 of the Gli1 protein and in place of Cys 515 of the Gli3T protein as described previously (Yuen et al., 2006). We previously showed that the insertion of inteins into these proteins at these positions abolishes their activities until splicing takes place (Yuen et al., 2006). The resulting constructs were introduced into HEK293 cells and splicing was evaluated by Western blot as described above.

Consistent with their enhanced performance characteristics in the GFP and mCherry contexts, the newly evolved inteins in Gli1 and in Gli3T resulted in significantly higher (~2- to 4-fold in Gli1, and ~3- to 8-fold in Gli3T) percentages of spliced protein compared with the 3-2 or 2-4 inteins (Figures 4B and 4C). Up to 48% and 60% of Gli1 protein was spliced by the newly evolved inteins after 12 and 24 hours, respectively, while the previously evolved 3-2 intein resulted in 22% and 32% splicing under the same conditions. Likewise, the newly evolved inteins generated up to 43% and 51% spliced Gli3T protein after 12 and 24 hours, compared with 10% and 18% for the 3-2 intein.

We previously noted (Yuen et al., 2006; unpublished data) that the 3-2 intein undergoes background splicing in the absence of 4-HT to a greater extent than the 2-4 intein, and the Gli1 and Gli3T data in Figures 4B and 4C replicated these observations. The background splicing of newly evolved clone 37R3-2 across all four proteins tested (GFP, mCherry, Gli1, and Gli3T) is very low and often was not detectable. Clone 30R3-1 in general resulted in a slightly higher degree of background splicing, but this level of splicing in the absence of 4-HT remained  $\leq 3\%$  in all four proteins tested in this work, and generally was similar to or slightly lower than the background splicing of the 3-2 intein. Taken together, these results establish the overall superior splicing kinetics and splicing efficiency without significant background splicing for both newly evolved inteins 37R3-2 and 30R3-1 in a variety of protein contexts in mammalian cells.

### Mutational Analysis of Evolved Inteins 30R3-1 and 37R3-2

To probe which mutations were responsible for the improved properties of evolved inteins 30R3-1 and 37R3-2, we systematically reverted each of the mutations (five in 30R3-1 and four in 37R3-2) to the corresponding amino acid present in the original 3-2 intein. Each of these reversion mutants was inserted genetically into GFP and transformed into RDY98 yeast cells. Protein expression was induced for 24 hours at 30 °C and the resulting cells were incubated for six hours in the presence or absence of 4-HT at 30 °C and 37 °C. Protein splicing was assessed by FACS and Western blot analysis as described above.

Compared with the 3-2 intein, both newly evolved intein variants contain Val34Ala and Thr328Lys mutations. These two changes also correspond to the two differences between the 2-4 intein and the 3-2 intein, which evolved from the 2-4 intein. In both newly evolved inteins the amino acids at these two positions were the residues present in the 2-4 intein. Consistent with the lower splicing activity in the absence or presence of 4-HT of the 2-4 intein relatively to the 3-2 intein, the reversion of Ala34 back to Val and Lys328 back to Thr resulted in significantly higher background splicing, especially in the 37R3-2 intein, together with slightly higher splicing efficiency in the presence of 4-HT (Figures 5A–5D). These results suggest that residues 34 and 328 in the newly evolved inteins modestly modulate splicing activity in a ligand-independent manner, and that the presence of Ala34 and Lys328 serves to decrease background splicing by a significant fraction (~2- to 4-fold), while lowering splicing efficiency in the presence of 4-HT by a much smaller relative fraction (~10% lower).

The Glu375Gly mutation was also present in both the 30R3-1 and the 37R3-2 inteins. Reversion of this mutation resulted in substantially increased (2- to 10-fold) background splicing in the absence of 4-HT without any significant change in splicing efficiency in the presence of 4-HT (Figures 5A–5D). This mutation therefore likely serves to suppress splicing activity in a manner that is selective for the conformation of the ligand-free intein.

Among the other two mutations in 30R3-1, neither was present in 37R3-2. Reverting the Thr at residue 66 to Ile in 30R3-1 resulted in no significant change in intein activity other than a slight decrease in 4-HT-triggered splicing at 30 °C (Figure 5A). Likewise, reversion of Pro124 to Leu also resulted in similar splicing activities in the presence or absence of 4-HT as the 30R3-1 intein (Figures 5A and 5B). These results suggest that Thr66 and Pro124 may not contribute to the observed changes in splicing activity, or may only contribute to improved splicing in combination with one or more additional mutations.

The Cys178Arg mutation is the sole change in 37R3-2 that is not present in 30R3-1. Reversion of this mutation modestly decreases splicing efficiency in the presence of 4-HT and may also slightly increase background splicing (Figures 5C and 5D). Interestingly, these results suggest that no single mutation in the 30R3-1 or 37R3-2 inteins is responsible for the substantial majority of the observed ~2- to 5-fold improved splicing of GFP in yeast at 30 °C or 37 °C compared with the parental 2-4 or 3-2 inteins. Instead, our observations suggest that the combination of four or five mutations, each of which contribute modest improvements in isolation, together result in the substantially faster and more efficient splicing in the presence of 4-HT while preserving or decreasing the extent of background splicing relative to the 3-2 intein.

## Discussion

The intein evolution efforts described here were performed under the hypothesis that inteins evolved under more stringent selection conditions and at 37 °C may yield ligand-dependent inteins with superior splicing yields, faster splicing kinetics, and/or lower background splicing than those previously reported. We used parallel 37 °C and 30 °C screening conditions to explore a wider range of possible advantageous mutations than might have been surveyed through 37 °C screens alone. Indeed, mutations in the two evolved inteins with the best overall properties, clones 30R3-1 and 37R3-2, arose from both the 30 °C and 37 °C libraries. These two inteins in the presence of 4-HT exhibited substantially higher yields of spliced protein and faster splicing, while maintaining comparable or slightly improved (decreased) amounts of background splicing in the absence of 4-HT.

It is interesting that the newly evolved inteins exhibited faster splicing kinetics (i.e., reached a higher percentage of final spliced protein levels at early time points) compared with the parental inteins even though the methods used in this work did not explicitly screen for improved splicing kinetics. It is possible that some of the mutations discovered in this work improve the kinetics of the splicing reaction itself, but it is equally likely that these mutations increased the rate of other steps in the ligand-induced splicing process such as protein folding, binding or dissociation from Hsp90 or other proteins (Feil et al., 1996; Kellendonk et al., 1996; Zhang et al., 1996; Danielian et al., 1998; Picard et al., 2000; Buskirk et al., 2004; Yuen et al., 2006), or conformational changes that influence the ability of the intein to undergo splicing. Because *in vitro* data on these ER-LBD containing inteins is currently not available due to the known difficulties of purifying the ER-LBD without fragmentation of the protein, mutagenesis, or binding to other stabilizing ligands (Nygaard and Harlow, 2001; Huang et al., 2006), more detailed mechanistic insights will likely require the biophysical analysis of the evolved inteins in living cells.



The FACS-based screening method used here, in which intein activity is coupled to an increase in GFP fluorescence, is ideally suited for this type of laboratory evolution in which starting proteins possessing detectable activities are evolved to higher levels of activity under specific sets of conditions. FACS offers a large dynamic range that is crucial for distinguishing active and highly active library members, allows analysis of individual library members at the single-cell level, and supports very high-throughput screens; in this work,  $\sim 10^7$  cells were screened in a few hours. FACS is also a nondestructive method, and yeast cells collected in this manner are robust enough to be cultured in liquid or on solid media immediately following the screening process. The ability to culture and thus amplify the cells resulting from each screen simplifies the process of enriching desired library members. These features together enabled improved variants to emerge by tuning the screen to capture progressively more fluorescent cells with progressively higher intein activity levels. Moreover, the fact that our reversion mutational analysis revealed that no single mutation resulted in most of the substantial improvements in splicing properties highlights the appropriateness of the function-based evolutionary approach taken in this study.

The use of small molecules to modulate protein structure post-translationally in living cells remains an attractive approach to studying protein function. The ligand-dependent intein, like other post-translationally triggered protein manipulation methods (Stankunas et al., 2003; Wang et al., 2003; Bayle et al., 2006), facilitates temporal control of protein structure as well as dose-dependent titration of spliced protein levels (Buskirk et al., 2004; Yuen et al., 2006). The intein offers some features that may make this tool particularly well suited for certain applications, especially given the improvements in splicing characteristics resulting from this work. While traditional chemical genetic approaches require the discovery of small molecules that perturb the activity of each protein of interest, the ligand-dependent intein confers dependence on a single small molecule (here, 4-HT) on a variety of proteins of interest with single-target specificity, albeit while requiring genetic modification. The protein splicing process leaves behind only a single Cys residue, or no scar in cases in which the target protein naturally contains a Cys residue in a location that results in loss of protein function upon intein insertion. Moreover, small-molecule-triggered protein splicing is pseudo-autocatalytic and does not require additional cellular components or specific conditions that may not be easy to establish for some intracellular proteins.

The small-molecule-dependent inteins developed here may be particularly suited for studying signaling pathways because of the minimal cellular perturbations required to achieve control over protein function. The use of the evolved inteins does not require changes to regulatory regions of genes and does not require the expression of any other proteins or nucleic acids. Since cell disruption is minimized, proteins that are a part of complex mammalian signaling pathways—for example those in which feedback regulation plays a significant role—have a greater chance of maintaining their native regulatory networks. Further, the dose-dependent nature of ligand-dependent intein-splicing allows for the fine control of functional protein levels. The newly evolved inteins, which in the presence of 4-HT generated  $\sim 40$ – $70\%$  of total expressed target protein as spliced protein in mammalian cells, are thus significantly better equipped than the first generation inteins to study protein function in cases in which the modest levels of spliced protein generated by the 2-4 and 3-2 inteins (typically  $10$ – $30\%$  of total expressed target protein) are insufficient to result in a meaningful physiological response.

## Significance

Elucidating the function of proteins in mammalian cells is particularly challenging due to the inherent complexity of these systems. Methods to study protein function in living cells ideally perturb the activity of only the protein of interest but otherwise maintain the natural

state of the host cell or organism. Small molecule-triggered inteins offer single-protein specificity and other desirable features as an approach to control protein function in cells post-translationally. The second-generation ligand-dependent inteins evolved in the present work splice to substantially higher yields and with faster kinetics in the presence of the cell-permeable small molecule 4-HT, especially at 37 °C, while exhibiting comparable or improved low levels of background splicing in the absence of 4-HT. These improvements were observed in all four protein contexts tested in mammalian cells at 37 °C, as well as in yeast cells assayed at 30 °C or 37 °C. The newly evolved inteins described in this work are therefore promising tools as conditional modulators of protein structure and function in yeast and mammalian cells.

## Experimental procedures

### Yeast Strains and Media

Media consisted of yeast nitrogen base (Sigma), 4% dextrose, and synthetic drop out supplements lacking uracil (MP Biomedical). Yeast were cultured in liquid medium or on agar plates at 30°C. The yeast strain RDY98 (*Erg6del::TRP1 pdr1del::KanMX pdr3::HIS3 ade2-1 trp1-1 his3-11,15 ura3-52 leu2-3,112 can1-100*) was provided by Professor Allen Buskirk at Brigham Young University. Protein induction was performed in media consisting of yeast nitrogen base (Sigma), 4% galactose, 4% raffinose, 0.4% dextrose, synthetic drop out supplements lacking uracil (MP Biomedical), and 1% of 100X penicillin-streptomycin solution (Cellgro) at 30 °C.

### Mammalian Cell Culture

HEK293 cells were cultured in Dulbecco's modified Eagle medium (DMEM):F12 medium with 10% fetal bovine serum (FBS) and 1% of 100X penicillin-streptomycin solution (Cellgro) according to standard protocols. Transient transfections were performed using Effectene (Qiagen) following the manufacturer's protocol.

### Protein Expression and 4-HT Treatment

All cell cultures were handled in parallel for each independent splicing efficiency experiment to ensure comparable levels of protein expression across the different samples.

Yeast cell cultures of the different intein constructs were started at the same time by inoculating colonies of transformed cells into the same preparation of yeast media and incubating the resulting cultures at 30 °C for 24 hours. In order to make cell growth phases more homogeneous,  $10^7$  cells from each of these cultures were incubated at the same time at 30 °C for an additional 24 hours. These cultures were then induced for protein expression at 30 °C in protein induction media for 24 hours. Following 24 hours of induction, a sample of  $10^7$  cells from each of the cultures was resuspended into protein induction media and incubated with or without 1  $\mu$ M 4-HT in 30 °C or 37 °C for the appropriate number of hours until harvesting for FACS or Western blot analysis. Yeast cells were harvested in stationary phase.

Mammalian cells were prepared by plating the same number of cells for transfection from the same parent plate to ensure all cells had been passaged at the same time and also the same number of times. After 24 hours in 37 °C following plating, these cells were transfected with the appropriate intein construct at the same time, then incubated at 37 °C for 24 hours. Following 24 hours, all of the plates were washed once with 37 °C PBS. The media was replaced with fresh media with or without 1  $\mu$ M 4-HT and incubated for 12 hours or 24 hours until harvesting for FACS or Western blot analysis.

## Library Construction

Error-prone PCR was carried out using 2-4 and 3-2 intein sequences as templates using DNA bases 8-oxo-2'-deoxyguanosine (8-oxo-dGTP) and 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP) purchased from TriLink BioTechnologies as previously described (Zaccolo et al., 1996) using oligonucleotides 5'-TAT GTA CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC GCA TGC-3' and 5'-GTG CAC GAC AAC CCC TTC GGC GAC GAG GGT GTG CAG TTC CTC GAC CTC GAG-3'. Mutagenized intein PCR products were inserted into p416Gal1 GFP-intein vector pre-cut with *SphI* and *XhoI* (to remove the existing intein sequence) by *in vivo* homologous recombination of overlapping PCR fragments as previously described (Raymond et al., 1999).

## Plasmid Construction

GFP-intein library members were amplified by PCR from the corresponding p416Gal1 library vector using oligonucleotides 5'-CTC GTT TAGTGA ACC GTC AGA GCC GCC ATG GCA AGC AAA GGA GAA -3 and 5'-CTA CTT GTC ATC GTC GTC CTT GTA ATC TTT GTA GAG CTC ATC CAT-3'. pFLAG-CMV-5.1 (Sigma) was amplified by PCR using oligonucleotides 5'-ACA CAT GGC ATG GAT GAG CTC TAC AAA GAT TAC AAG GAC GAC GAT-3' and 5'-TTC TTC TCC TTT GCT TGC CAT GGC GGCTCT GAC GGTTCA CTA AAC-3'. The PCR products were ligated together through isothermal assembly (Gibson et al., 2009). The resulting ligated vectors were purified using Min-Elute columns (Qiagen) and eluted with 10 µL deionized water. 1 µL of this elution was transformed into NEB Turbo chemically competent *E. coli* cells (New England Biolabs) and plated onto LB + carbenicillin agar plates. Plates were incubated overnight at 37 °C, and individual colonies were picked and sequenced to verify correct plasmid construction.

p3XFlag-CMV-14 (Sigma) vectors with Gli1-intein (2-4 and 3-2), and Gli3T-intein (2-4 and 3-2) sequences were previously described (Yuen et al., 2006). p3XFlag-CMV-14 mCherry-intein (2-4 and 3-2) vectors were constructed by cloning the 2-4 and 3-2 intein sequences into the pRSET-B mCherry vector provided by Professor Roger Tsien (University of California at San Diego). The mCherry-intein 2-4 and mCherry-intein 3-2 sequences were then amplified using oligonucleotides introducing a 5' *EcoRI* site and a 3' *BamHI* site and ligated into *EcoRI*- and *BamHI*-digested p3XFlag-CMV-14 vector. 30R3-1 intein sequence and 37R3-2 intein sequence were amplified using the following oligonucleotides compatible with mCherry, Gli1, or Gli3 contexts.

mCherry: 5'-TTC GAG GAC GGC GGC GTG GTG ACC GTG TGC CTT GCC GAG GGT ACC-3' and 5'-GCC GTC CTG CAG GGA GGA GTC CTG GCA GTT GTG CAC GAC AAC CCC-3'

Gli1: 5'-ATC CAC GGG GAG CGG AAG GAA TTC GTG TGC CTT GCC GAG GGT ACC-3' and 5'-CTC CCT CGA GCA ACC TCC CCA ATG GCA GTT GTG CAC GAC AAC CCC-3'

Gli3: 5'-ATT CAT GGA GAA AAG AAG GAA TTC GTG TGC CTT GCC GAG GGT ACC-3' and 5'-CTC TCG AGA ACA ATC AAG CCA GCG GCA GTT GTG CAC GAC AAC CCC-3'

The p3XFlag-pCMV-14 vector was amplified using oligonucleotides 5'-CTC GTC GCC GAA GGG GTT GTC-3' and 5'-ATC GAA GAT TCG GGT ACC CTC-3'. The intein PCR products and the vector PCR product were ligated together through the isothermal assembly method (Gibson et al., 2009) and the resulting ligated material was treated as discussed above.

## FACS Screening and Analysis

Yeast cells transformed with library plasmids were cultured for 24 hrs in the appropriate synthetic drop out media in 30 °C. Cells were washed and resuspended in protein induction media and cultured for another 24 hrs at 30 °C. After 24 hrs of protein induction, cells were treated with 1  $\mu$ M 4-HT or left untreated with 4-HT as appropriate for the prescribed amount of time in either 30 °C or 37 °C. After the appropriate length of time, cells were harvested by washing once in PBS, then resuspended in PBS with 0.1% bovine serum albumin (BSA) (Sigma). Cell sorting was performed using a MoFlo cell sorter (DakoCytomation). Cell fluorescence analysis was carried out on a BD LSRII cell analyzer.

HEK293 cells were grown in 10 cm dishes or 6-well plates and transfected with relevant mammalian vectors using Effectene. After growth in the absence of 4-HT or in the presence of 1  $\mu$ M 4-HT for 24 hours, cells were trypsinized and resuspended in 500  $\mu$ L of phosphate buffered saline with 1% FBS and 75 U/mL DNase (New England Biolabs). Cell fluorescence analysis was carried out on a BD LSRII cell analyzer.

## Western Blots

Western blots were performed using Nu-PAGE 12% Bis-Tris gels (Invitrogen) in 3-(*N*-morpholino)propanesulfonic acid (MOPS)-sodium dodecyl sulfate (SDS) buffer (Invitrogen). SDS-polyacrylamide gel electrophoresis (PAGE) and Western blotting were performed using standard protocols. Gels were transferred onto polyvinylidene fluoride (PDVF) membranes (Millipore). Western blots were processed using a mouse anti-FLAG antibody (Sigma) as the primary antibody and a secondary Alexa Fluor 800-conjugated goat anti-mouse antibody (Li-cor Biosciences), then visualized and quantitated using an Odyssey imager (Li-cor Biosciences).

## Reversion Mutant Construction

Each evolved amino acid change was the result of a single nucleotide mutation. Each reversion mutant was generated using the QuikChange method (Stratagene) with Pfu Turbo and the following oligonucleotides (the mutated base pair is underlined in each oligonucleotide pair).

A34V: 5'-CGC AAG CCT ATT CAT GTC GTG GCT GTT GCC AAG GAC GGA ACG CTG CTC GCG-3' and 5'-CGC GAG CAG CGT TCC GTC CTT GGC AAC AGC CAC GAC ATG AAT AGG CTT GCG-3'

T66I: 5'-GGG TTG CGG ATC GCC GGT GGC GCC ATC GTG TGG GCG ACA CCC GAT CAC AAG-3' and 5'-CTT GTG ATC GGG TGT CGC CCA CAC GAT GGC GCC ACC GGC GAT CCG CAA CCC-3'

P124L: 5'-TTG TTG GAT GCT GAG CCC CCC ATA CTC TAT TCC GAG TAT GAT CCT ACC AGT-3' and 5'-ACT GGT AGG ATC ATA CTC GGA ATA GAG TAT GGG GGG CTC AGC ATC CAA CAA-3'

C178R: 5'-CCA TGA TCA GGC CCA CCT TCT AGA ACG TGC CTG GCT AGA GAT CCT GAT GAT-3' and 5'-ATC ATC AGG ATC TCT AGC CAG GCA CGT TCT AGA AGG TGG GCC TGA TCA TGG-3'

K328T: 5'-GAG CAT CTG TAC AGC ATG AAG TAC ACG AAC GTG GTG CCC CTC TAT GAC CTG-3' and 5'-CAG GTC ATA GAG GGG CAC CAC GTT CGT GTA CTT CAT GCT GTA CAG ATG CTC-3'

G375E: 5'-TTC CTG CAC GAC ATG CTG GCG GAA GAA CTC CGC TAT TCC GTG  
ATC CGA GAA-3' and 5'-TTC TCG GAT CAC GGA ATA GCG GAG TTC TTC CGC  
CAG CAT GTC GTG CAG GAA-3'

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by NIH/NIGMS R01 GM065400 and by the Howard Hughes Medical Institute. I.C. was supported by an NIH/NRSA Postdoctoral Fellowship and by DARPA HR0011-08-0085.

## References

- Acar M, Pando BF, Arnold FH, Elowitz MB, van Oudenaarden A. A general mechanism for network-dosage compensation in gene circuits. *Science*. 2010; 329:1656–1660. [PubMed: 20929850]
- Bai CB, Stephen D, Joyner AL. All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. *Dev Cell*. 2004; 6:103–115. [PubMed: 14723851]
- Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*. 2006; 126:995–1004. [PubMed: 16959577]
- Banaszynski LA, Wandless TJ. Conditional control of protein function. *Chem Biol*. 2006; 13:11–21. [PubMed: 16426967]
- Barna M, Pandolfi PP, Niswander L. Gli3 and Plzf cooperate in proximal limb patterning at early stages of limb development. *Nature*. 2005; 436:277–281. [PubMed: 16015334]
- Bayle JH, Grimley JS, Stankunas K, Gestwicki JE, Wandless TJ, Crabtree GR. Rapamycin analogs with differential binding specificity permit orthogonal control of protein activity. *Chem Biol*. 2006; 13:99–107. [PubMed: 16426976]
- Buskirk AR, Liu DR. Creating small-molecule-dependent switches to modulate biological functions. *Chem Biol*. 2005; 12:151–161. [PubMed: 15734643]
- Buskirk AR, Ong YC, Gartner ZJ, Liu DR. Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc Natl Acad Sci U S A*. 2004; 101:10505–10510. [PubMed: 15247421]
- Danielian PS, Muccino D, Rowitch DH, Michael SK, McMahon AP. Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr Biol*. 1998; 8:1323–1326. [PubMed: 9843687]
- Feil R, Brocard J, Mascres B, LeMeur M, Metzger D, Chambon P. Ligand-activated site-specific recombination in mice. *Proc Natl Acad Sci U S A*. 1996; 93:10887–10890. [PubMed: 8855277]
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998; 391:806–811. [PubMed: 9486653]
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009; 6:343–345. [PubMed: 19363495]
- Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A*. 1992; 89:5547–5551. [PubMed: 1319065]
- Hartley PD, Madhani HD. Mechanisms that specify promoter nucleosome location and identity. *Cell*. 2009; 137:445–458. [PubMed: 19410542]
- Huang J, Koide A, Nettle KW, Greene GL, Koide S. Conformation-specific affinity purification of proteins using engineered binding proteins: application to the estrogen receptor. *Protein Expr Purif*. 2006; 47:348–354. [PubMed: 16325424]

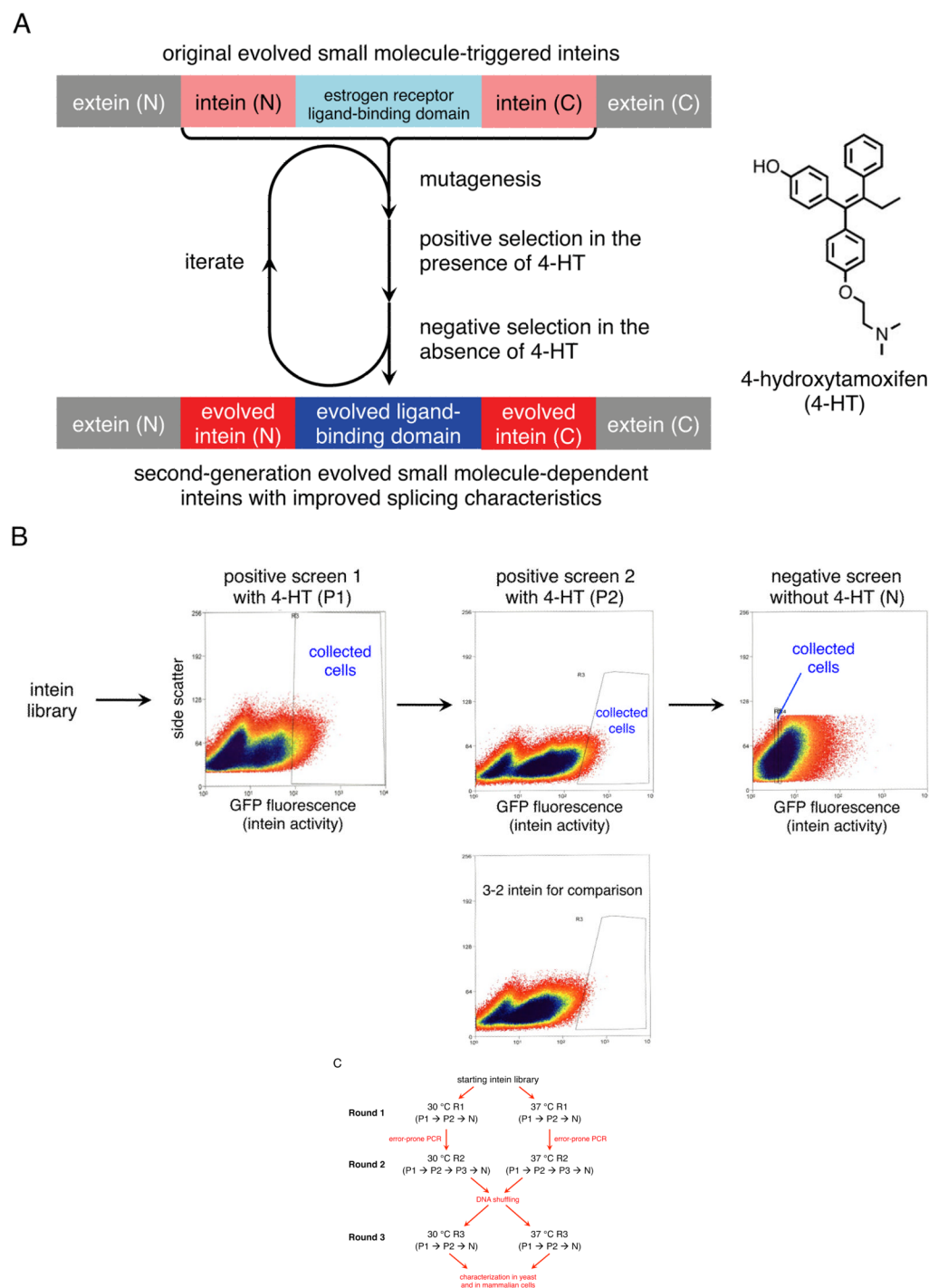


- Kellendonk C, Tronche F, Monaghan AP, Angrand PO, Stewart F, Schutz G. Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res.* 1996; 24:1404–1411. [PubMed: 8628671]
- Koebernick K, Pieler T. Gli-type zinc finger proteins as bipotential transducers of Hedgehog signaling. *Differentiation.* 2002; 70:69–76. [PubMed: 12076333]
- Lew BM, Paulus H. An in vivo screening system against protein splicing useful for the isolation of non-splicing mutants or inhibitors of the RecA intein of *Mycobacterium tuberculosis*. *Gene.* 2002; 282:169–177. [PubMed: 11814689]
- Marschang P, Brich J, Weeber EJ, Sweatt JD, Shelton JM, Richardson JA, Hammer RE, Herz J. Normal development and fertility of knockout mice lacking the tumor suppressor gene LRP1b suggest functional compensation by LRP1. *Mol Cell Biol.* 2004; 24:3782–3793. [PubMed: 15082773]
- Mootz HD, Blum ES, Muir TW. Activation of an autoregulated protein kinase by conditional protein splicing. *Angew Chem Int Ed Engl.* 2004; 43:5189–5192. [PubMed: 15455415]
- Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW. Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J Am Chem Soc.* 2003; 125:10561–10569. [PubMed: 12940738]
- Mootz HD, Muir TW. Protein splicing triggered by a small molecule. *J Am Chem Soc.* 2002; 124:9044–9045. [PubMed: 12148996]
- Nygaard FB, Harlow KW. Heterologous expression of soluble, active proteins in *Escherichia coli*: the human estrogen receptor hormone-binding domain as paradigm. *Protein Expr Purif.* 2001; 21:500–509. [PubMed: 11281726]
- Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science.* 1996; 273:1392–1395. [PubMed: 8703075]
- Paulus H. Protein splicing and related forms of protein autoprocesing. *Annu Rev Biochem.* 2000; 69:447–496. [PubMed: 10966466]
- Picard D. Posttranslational regulation of proteins by fusions to steroid-binding domains. *Methods Enzymol.* 2000; 327:385–401. [PubMed: 11044998]
- Pratt MR, Schwartz EC, Muir TW. Small-molecule-mediated rescue of protein function by an inducible proteolytic shunt. *Proc Natl Acad Sci U S A.* 2007; 104:11209–11214. [PubMed: 17563385]
- Raymond CK, Pownder TA, Sexson SL. General method for plasmid construction using homologous recombination. *Biotechniques.* 1999; 26:134–138. 140–131. [PubMed: 9894602]
- Sauer B, Henderson N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A.* 1988; 85:5166–5170. [PubMed: 2839833]
- Schneekloth JS Jr, Fonseca FN, Koldobskiy M, Mandal A, Deshaies R, Sakamoto K, Crews CM. Chemical genetic control of protein levels: selective in vivo targeted degradation. *J Am Chem Soc.* 2004; 126:3748–3754. [PubMed: 15038727]
- Schreiber SL. The small-molecule approach to biology: Chemical genetics and diversity-oriented organic synthesis make possible the systematic exploration of biology. *Chem Eng News.* 2003:51–61.
- Schwartz EC, Saez L, Young MW, Muir TW. Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol.* 2007; 3:50–54. [PubMed: 17128262]
- Shi J, Muir TW. Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J Am Chem Soc.* 2005; 127:6198–6206. [PubMed: 15853324]
- Shogren-Knaak MA, Alaimo PJ, Shokat KM. Recent advances in chemical approaches to the study of biological systems. *Annu Rev Cell Dev Biol.* 2001; 17:405–433. [PubMed: 11687495]
- Specht KM, Shokat KM. The emerging power of chemical genetics. *Curr Opin Cell Biol.* 2002; 14:155–159. [PubMed: 11891113]
- Stankunas K, Bayle JH, Gestwicki JE, Lin YM, Wandless TJ, Crabtree GR. Conditional protein alleles using knockin mice and a chemical inducer of dimerization. *Mol Cell.* 2003; 12:1615–1624. [PubMed: 14690613]

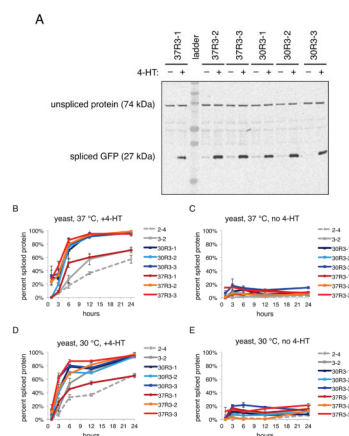
- Wang B, Fallon JF, Beachy PA. Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell*. 2000; 100:423–434. [PubMed: 10693759]
- Wang H, Shimizu E, Tang YP, Cho M, Kyin M, Zuo W, Robinson DA, Alaimo PJ, Zhang C, Morimoto H, et al. Inducible protein knockout reveals temporal requirement of CaMKII reactivation for memory consolidation in the brain. *Proc Natl Acad Sci U S A*. 2003; 100:4287–4292. [PubMed: 12646704]
- Wong SL, Roth FP. Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in *Saccharomyces cerevisiae*. *Genetics*. 2005; 171:829–833. [PubMed: 15998714]
- Xu MQ, Southworth MW, Mersha FB, Hornstra LJ, Perler FB. In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell*. 1993; 75:1371–1377. [PubMed: 8269515]
- Yuen CM, Rodda SJ, Vokes SA, McMahon AP, Liu DR. Control of transcription factor activity and osteoblast differentiation in mammalian cells using an evolved small-molecule-dependent intein. *J Am Chem Soc*. 2006; 128:8939–8946. [PubMed: 16819890]
- Zaccolo M, Williams DM, Brown DM, Gherardi E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol*. 1996; 255:589–603. [PubMed: 8568899]
- Zhang Y, Riesterer C, Ayrall AM, Sablitzky F, Littlewood TD, Reth M. Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res*. 1996; 24:543–548. [PubMed: 8604292]
- Zhao H, Zha W. In vitro ‘sexual’ evolution through the PCR-based staggered extension process (StEP). *Nat Protoc*. 2006; 1:1865–1871. [PubMed: 17487170]

### Highlights

- Small-molecule-dependent inteins enable protein structure and function to be controlled post-translationally in living cells.
- Starting with two previously evolved ligand-dependent inteins that splice in a manner dependent on 4-hydroxytamoxifen (4-HT), we evolved second-generation inteins with substantially improved splicing efficiency in both yeast and mammalian cells.
- The newly evolved inteins were tested in four different protein contexts in yeast and human cells and typically resulted in 50–90% spliced protein in the presence of 4-HT and < 5% splicing in the absence of 4-HT.

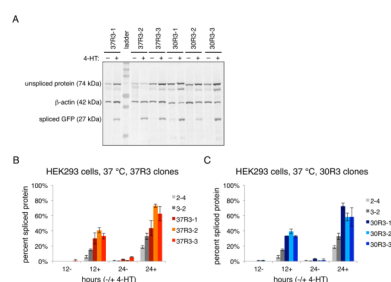
**Figure 1.**

Intein evolution approach. (A) Overview of the directed evolution strategy used to isolate improved small molecule-dependent inteins. (B) Each round of evolution consisted of mutagenesis followed by at least two positive FACS screens in the presence of 1  $\mu$ M 4-HT and one negative FACS screen in the absence of 4-HT. One set of FACS data from the Round 1 positive and negative screens is shown. (C) Two intein evolution efforts were performed in parallel at 30 °C and 37 °C, comprising 20 total screening steps.

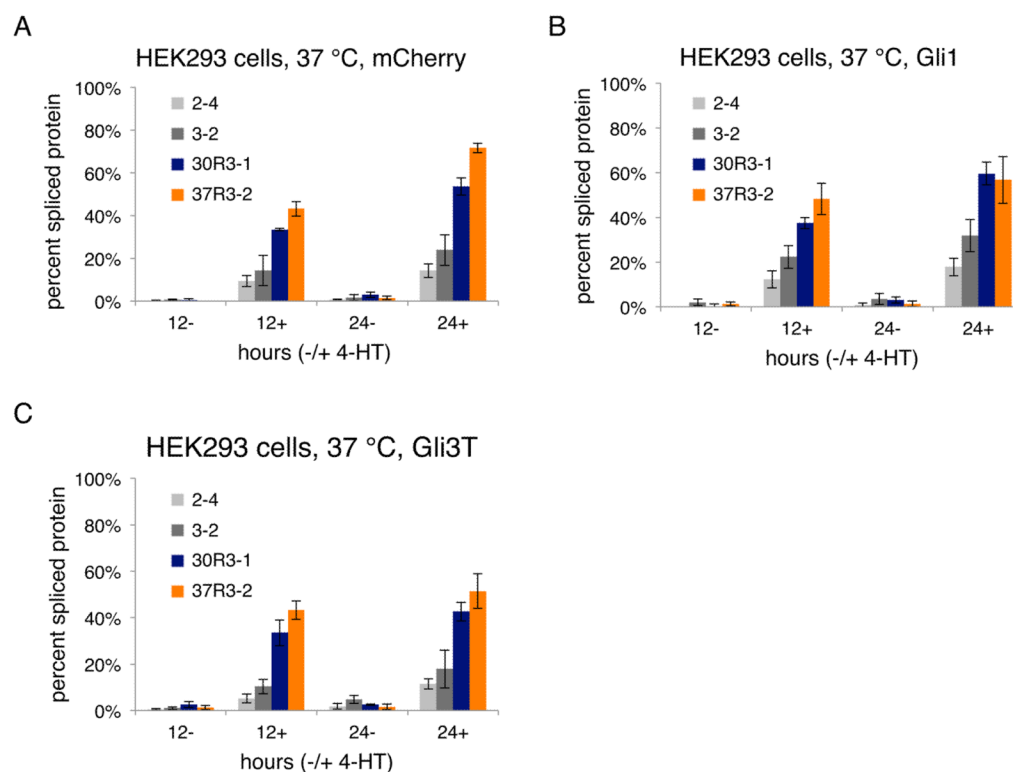


**Figure 2.** Characterization of newly evolved inteins in yeast cells. (A) Representative Western blot of lysates from yeast cells expressing evolved intein variants in the context of GFP. Each lane shows lysate from  $2.5 \times 10^6$  cells after six hours of growth at 30 °C in the absence or presence of 1  $\mu$ M 4-HT visualized with an anti-FLAG-tag antibody. Quantitation of spliced and unspliced protein bands by densitometry was used to calculate the percent spliced protein shown in the rest of the figure. (B)–(E) Inteins splicing characteristics in yeast at various time points in the context of GFP at 30 °C ((D) and (E)) or 37 °C ((B) and (C)), either with ((B) and (D)) or without ((C) and (E)) 1  $\mu$ M 4-HT. Error bars represent the standard deviation of at least three independent experiments. (See also Figure S1.)

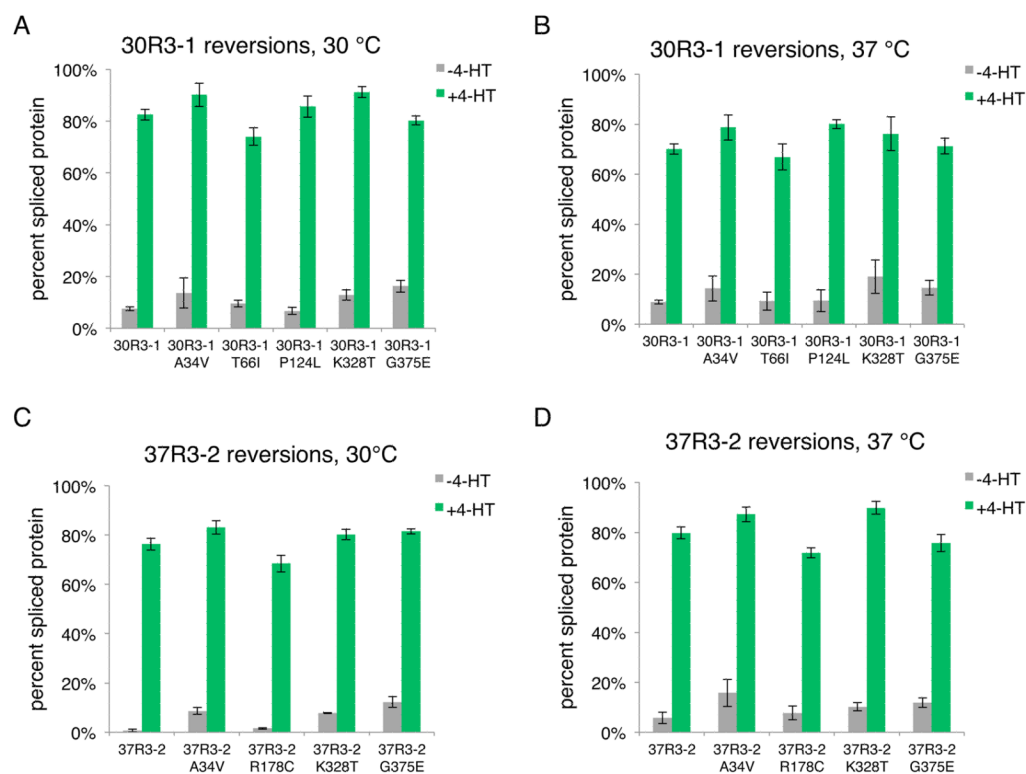


**Figure 3.**

Characterization of newly evolved inteins in mammalian cells. (A) Representative Western blot of lysates from HEK293 cells expressing evolved intein variants in the context of GFP. Each lane shows lysate from cells after 12 hours of growth at 37 °C in the absence or presence of 1  $\mu$ M 4-HT processed with an anti-FLAG-tag antibody to visualize spliced and unspliced GFP, and an anti- $\beta$ -actin antibody to visualize  $\beta$ -actin, which served as a loading control. Quantitation of spliced and unspliced protein bands by densitometry was used to calculate the percent spliced protein shown in the rest of the figure. (B) and (C) Splicing characteristics of inteins in the GFP context in HEK293 cells at 37 °C after 12 and 24 hours incubation in the presence or absence of 1  $\mu$ M 4-HT. Three evolved inteins from the 37 °C evolution effort are shown in (B), and three evolved inteins from the 30 °C evolution effort are shown in (C). Error bars represent the standard deviation of at least three independent experiments. (See also Figure S2.)

**Figure 4.**

Splicing characteristics of the 30R3-1 and 37R3-2 evolved inteins in mammalian cells in three different protein contexts. HEK293 cells expressing the inteins shown in the context of mCherry (A) (See also Figure S3), Gli1 (B), or Gli3T (C) were incubated for 12 h or 24 h in the presence or absence of 1  $\mu$ M 4-HT. Unspliced and spliced protein was quantitated as described in Figure 3. Error bars represent the standard deviation of at least three independent experiments.

**Figure 5.**

Reversion mutant analysis of evolved inteins 30R3-1 and 37R3-2. Each mutation in 30R3-1 ((A) and (B)) and 37R3-2 ((C) and (D)) relative to the original 3-2 intein was reverted separately and the resulting intein variants in the context of GFP were characterized in yeast cells at 30 °C ((A) and (C)) and at 37 °C ((B) and (D)). Yeast cell lysates were prepared and analyzed by Western blot and densitometry after 6 hours as described in Figure 2. Error bars represent the standard deviation of at least three independent experiments.

Mutations isolated in evolved inteins. Three clones each from the 30 °C and 37 °C evolution efforts were chosen based on their abundance among DNA sequences surviving Round 3. The mutations compared with the 3-2 intein sequence are shown. Mutations in the intein are shown in red; those in the ligand-binding domain are shown in blue.

Table 1

evolved intein clone	intein mutations		ligand-binding domain mutations			intein mutation
30R3-1	Val34Ala	Ile66Thr	Leu124Pro		Thr328Lys	Glu375Gly
30R3-2	Val34Ala	Ile66Thr			Thr328Lys	
30R3-3	Val34Ala	Ile66Thr	Leu124Pro		Thr328Lys	
37R3-1	Val34Ala	Ile66Thr		Asp129Asn	Thr328Lys	Glu375Gly
37R3-2	Val34Ala	Ile66Thr			Thr328Lys	
37R3-3		Ile66Thr			Thr328Lys	